Understanding Usages by Modeling Diversity over Time

Amaury L'Huillier, Sylvain Castagnos, and Anne Boyer

KIWI Team, LORIA - Université de Lorraine
Campus Scientifique, B.P. 239
54506 Vandœuvre - France
{amaury.lhuillier,sylvain.castagnos,anne.boyer}@loria.fr

Abstract. Let's imagine a system that can recommend the kind of music (among other application domains) you like to listen when you are at work, without having to know your location, IP address or even to ask your current mood. In this paper, we bring this dream closer by proposing a model that can automatically understand the user's current context. This model, called DANCE, analyzes the attributes of the items in your recent history and monitors the relative diversity brought by your consultations over time. We validated our approach with a music corpus of 100 users and a global history of 204,758 plays.

Keywords: User Modeling, Analysis of usages, Diversity, Recommenders.

1 Introduction

Recommender systems aim at improving human-computer interactions between the general audience and online services such as e-commerce websites, VOD, music streaming or search engines. These systems rely on various machine learning techniques to understand users' behaviors, to build preference models, and to predict future interests or needs so as to personalize information access. After two decades of researches and uses in industry, recommender systems have been proven to be useful both for companies (for knowing their customers, increasing revenues) and users (gain of time, usefulness of recommendations). However, most of recommenders only focus on the precision of the recommendations, now reaching excellent performances. They do not take enough into account human factors involved in the decision making process, such as context, confidence, explanations and need for diversity. These factors are yet crucial to maximize users' satisfaction, as highlighted by Jones in [7]: while a difference of 10% of the RMSE between two algorithms cannot be perceived by users, recommendations made at the right time in the proper way can double users' acceptance rate.

Our work-in-progress follows on from the experiment we conducted in [3], which aimed at identifying the steps of a user's decision process, when facing a recommender system on a e-commerce website. In that study, the diversity of recommendations has been shown to be correlated with the active user's confidence level and his intention to buy. Our user study strongly suggested to favor approaches where similarity between recommendations progressively increases at the product brokering stage, and where diversity is re-introduced in the end of the session (when decision is close to being made). However, as far as we know, no temporal model has ever been proposed in the literature to adapt recommendations to the active user's evolving need for diversity over time.

As a first step towards this objective, we propose a new way to analyze usages by modeling the diversity brought by each consulted item relatively to a short user's history. We took inspiration from natural language processing techniques [1] to survey the diversity over time within a sequence of consultations, and introduce generic multi-attribute diversity metrics. In that way, we can automatically understand the current context of the active user in real time. We called our model DANCE, which is the acronym for "Diversity And Natural Context Elicitation". DANCE offers the advantage of being highly scalable and generic, preserves privacy and does not require an ontology to put words on the context. We validated our approach on a music corpus and have been able to detect skipped songs and ends of sessions within users' listening sequences.

This paper is organized as follows: Section 2 is an overview of the state of the art of diversity in recommender systems. Section 3 is dedicated to the presentation of our model and Section 4 presents and discusses its performances.

2 Related Work

2.1 Usages of diversity

Recommender systems usually rely on similarity measures between users' preferences, contents of items or past usages to predict future interests as accurately as possible, even if accuracy is insufficient to maximize user's satisfaction [9]. Several recent research papers encourage the use of diversity, in addition to classic evaluation metrics, to measure the quality of algorithms so as to discredit those focusing on certain parts of the item spectrum [6].

The diversity has been defined by Smyth and McClave [11] as the opposite dimension to similarity. More recent works refined this definition as the measure that quantifies the dissimilarity within a set of items [4]. McGinty and Smyth [8] have been the first to show that diversity improves the efficiency of recommendations. They see diversity as a contribution in response to a bad recommendation, to provide new exploration strategies. Other usages include intrinsic diversity which avoids redundancy between the items to be recommended [5], and extrinsic diversity which aims at alleviating the uncertainty due to data ambiguity or sparsity in user preference models by recommending a large set of items [10]. Nevertheless, all these works try to manage diversity at a given moment, but do not integer the time dimension.

In [3], we aimed at identifying the different steps that mark out the decision-making process to get deeper into details, and pay attention to why and when diversity should be provided. It comes out that the user's need for diversity evolves over time, and is particularly high when the user is close to reach a decision or to change his mind. Starting from this conclusion, the core idea of this paper consists in modeling the evolution of diversity over time within sequences of consultations so as to automatically detect and understand the user context. Time-aware and context-aware recommender systems are receiving increasing attention, since it has been proven to be an effective approach in the Netflix Prize competition [2]. However, current context-aware systems often need an *a priori* knowledge of various possible contexts (work vs. home, holidays or not, season, moment of the day, etc.), with various representations such as ontologies, to filter information. Our long-term goal is to detect changes of context and common features/patterns inside these contexts thanks to diversity, without any prior context knowledge nor user involvement.

2.2 Metrics of diversity

In the literature, most of diversity metrics rely on the similarity between items: the more the items are similar, the lower is the diversity between them [11]. Similarity between two items is defined as the weighted sum of the similarities on their attributes (see (1)). A is the set of attributes, with $card(A) = n$.

$$sim_A(i_1,i_2) = \frac{\sum_{j=1..n} w_j * sim_{attribute=j}(i_1,i_2)}{\sum_{j=1..n} w_j} \tag{1}$$

Starting from this similarity metric, Ziegler *et al.* defined an Intra-List Similarity metric (ILS) which computes the average similarity – and by opposition the average diversity – within a class a class C, made up of m items. Smyth and McClave [11] has introduced the same kind of metric, called $Diversity$, which computes the average dissimilarity within the class C. Additionally, they define the notion of relative diversity (*RelDiversity*) as the added value in terms of diversity of an item i on a class of items C for a user u, while considering all items' attributes simultaneously (see (2)).

$$RD^u(i,C) = \begin{cases} 0 \text{ if C} = \{\}, \\ \frac{\sum_{j=1..m}(1-sim_A(i,c_j))}{m} \text{ otherwise.} \end{cases} \tag{2}$$

3 Our Model

As explained above, our model intends to analyze usages so as to detect the changes of user context. We called this model DANCE, acronym for "Diversity And Natural Context Elicitation", since it took inspiration from natural language processing techniques and focus on diversity. We consider the active user's history of consultations as a contiguous sequence of items. As in a k-order Markov model, we assume that the consultation of an item i only depends on the k previously accessed items. These k items are called *history*. We then measure the relative diversity brought by i relatively to the k previous consulted items. Our model can compute a global score of relative diversity by taking all the attributes of items into account together, or can monitor the evolution of relative diversity for each attribute separately.

Within the framework of this model, we define the notion of *context* as a period of time during which consulted items share common features. During this interval, items have similar values for one or several attributes, which means that the relative diversity evolve slowly over time for these attributes (and potentially for the global relative diversity). Otherwise, when the relative diversity of the current item increases more than a given threshold, the user context is about to change. Consequently, in order to detect these changes of context, our model looks for local maximums on the curve of relative diversity (i.e. where there is an horizontal tangent after a significant increase of the relative diversity in comparison with previous time steps).

In the following, we note $I = \{i_1, i_2, ..., i_m\}$ the global set of items available for the active user. Each item has a fixed number of attributes ($card(A) = n$). The set of items actually consulted by the active user u is $C_u = \{c_1^u, ..., c_k^u\}$, but we do not need to remember it since we only observe the recent *history* of size k. The latter can be decomposed under the form of a tuple, written $< c_{t-k}^u, ..., c_{t-2}^u, c_{t-1}^u >$, where t is the current time step. At last, the value of the attribute a for the item c_t^u is written $c_t^u.a$.

The DANCE model consists in computing the relative diversity brought by c_t^u in comparison with the *history*, using the formula (2) above at each time step t, that is to say: $RD^u(c_t^u, < c_{t-k}^u, ..., c_{t-2}^u, c_{t-1}^u >)$. The function of relative diversity RD^u, and at the same time the function of similarity (see (1)) requires to define sub-functions of similarity per attribute.

If the attribute a can take values in (R), this sub-function is computed using Equation (3), where min_a and max_a respectively correspond to minimum and maximum values of a on the set of items I.

$$sim_{attribute=a}(c_t^u, c_{t-1}^u) = e^{-10*\left(\frac{c_t^u.a - c_{t-1}^u.a}{max_a - mina}\right)^2} \tag{3}$$

If the attribute a contains a list of terms, the sub-function of similarity per attribute is given by Equation (4).

$$sim_{attribute=a}(c_t^u, c_{t-1}^u) = \frac{card(c_t^u.a \cap c_{t-1}^u.a)}{min(card(c_t^u.a), card(c_{t-1}^u.a))} \tag{4}$$

For a given user and a fixed size of *history*, our algorithm has a complexity in constant time. It can perform computations in real time, while preserving his privacy since we only need his short *history* and attribute values of items.

4 Performance Analysis

We based our evaluation on a music corpus, due to its numerous advantages. First, music items can be characterized by a large number of attributes and information about the songs are easy to retrieve. Besides, even though very similar to analyses of sequences in other application domains (e-commerce websites, VOD), it avoids some bias. Songs are short enough to be fully listened in a sequence contrary to movies, and we can record list of songs that have been skipped before the end. The duration of a consultation is independent from the reading speed of the subject, by opposition to textual websites. At last, users of music streaming services like to discover new songs.

We used the API of Last.fm[1] to collect data of 100 users for a period between June 2005 and October 2013. For each user, we own a set of music plays which contains the names of the songs, the names of the artists and timestamps corresponding to the times he listened each song. We extracted values of attributes about these songs using the Echonest[2] API. Information about the various item attributes and possible associated values are summarized in Table (1).

	music							artist	
attribute	duration	tempo	mode	loudness	energie	hotttness	danceability	hotttness	familiarity
max	4194	239	1	41.76	0.99	0.91	0.98	0.99	0.91
min	12	0	0	0.08	0.00002	0.000782	0.039	0.109	0.051
average	218.09	128.25	0.57	7.68	0.76	0.33	0.43	0.62	0.63
deviation	84.25	30.31	0.49	4.19	0.21	0.12	0.17	0.12	0.13

Table 1. Characteristics of the corpus.

To test the ability of our model to detect the changes of context in users' sequences, we decided to mark the ends of sessions for each user, and to record the list of songs that have been skipped. In order to do so, we fixed the maximum duration without any music play before considering a session as ended to 15 minutes. We obtained 18,640 sessions composed of 204,758 plays (40,923 distinct songs and 5,571 distinct artists), with an average of 10.98 songs per session (39.92 minutes). After that, we put the whole list of plays for each user in a single sequence (test data), and tried to retrieve these changes of context (see Table (2). We tested our model with four-gram (i.e. with *history* size $k = 3$).

	Test data	Changes detected	% of detections
Ends of sessions	18,640	13,551	74.30 %
Number of skipped songs	28,849	2,752	9.53 %
Total number of local maximums in RD^u		25,193	

Table 2. Detecting the changes of context with the DANCE model.

Taking in consideration all the attributes with the same weight, we were able to retrieve 74.30% of the ends of sessions contained in our dataset. Moreover, only 9.53% of the skipped songs were detected. These results are not very surprising, since we are more likely to detect differences between sessions (making them

[1] `www.lastfm.fr/`

[2] `http://developer.echonest.com/`

more identifiable), rather than to detect anomalies with skipped songs: a user can skip a song even if this one is not diverse (e.g. because he is fed up).

5 Conclusion and Perspectives

In this paper, we showed that modeling diversity over time is an elegant and efficient way to detect users' changes of context when analyzing usages in real time. The biggest strength of our model is to be able to automatically detect similarities within contexts, and dissimilarities between contexts. In Table (2), we can notice that there are more local maximums obtained with RD^u (25,193 peaks of diversity) than the number of sessions. This means that, inside a session, we can find several contexts and take advantage of them. These preliminary results constitute a first step to adapt recommender systems to users' context, by explaining recommendation with common session features, and by varying the level of diversity to users' needs and expectations. In the same time, we plan to optimize the weights of attributes and size of history.

References

1. Brun, A., Bonnin, G., Boyer, A.: History Dependent Recommender Systems based on Partial Matching. In: UMAP, pp. 343–348 (2009)
2. Campos, P.G., Díez, F., Cantador, I.: Time-Aware Recommender Systems: A Comprehensive Survey and Analysis of Existing Evaluation Protocols. Journal on User Modeling and User-Adapted Interaction, Volume 24, pp. 67–119 (2014)
3. Castagnos, S., Jones, N., Pu, P.: Eye-Tracking Product Recommenders' Usage. In: Proc. of the 4th ACM Conference on Recommender Systems, pp. 29–36 (2010)
4. Castagnos, S., Brun, A., Boyer, A.: When Diversity Is Needed... But Not Expected! In: IMMM, pp. 44–50 (2013)
5. Clarke, C.L.A. et al.: Novelty and Diversity in Information Retrieval Evaluation. In: Proc. of the 31st International ACM Conference on Research and Development in Information Retrieval, pp. 659–666 (2008)
6. Jannach, D., Lerche, L., Gedikli, F., Bonnin, G.: What Recommenders Recommend – An Analysis of Accuracy, Popularity, and Sales Diversity Effects. In: Proc. of User Modeling, Adaptation, and Personalization, pp. 25–37 (2013)
7. Jones, N.: User Perceived Qualities and Acceptance of Recommender Systems: The Role of Diversity. PhD Thesis, EPFL, Lausanne, Switzerland (2010)
8. McGinty, L., Smyth, B.: On the Role of Diversity in Conversational Recommender Systems. In: Int. Conf. on Case-Based Reasoning, pp. 276–290 (2003)
9. McNee, S.M., Riedl, J., Konstan, J.A.: Being Accurate Is Not Enough: How Accuracy Metrics Have Hurt Recommender Systems. In: Proc. of Computer-Human Interaction, pp. 1097–1101 (2006)
10. Radlinski, F. et al.: Redundancy, Diversity and Interdependent Document Relevance. Journal SIGIR Forum, Volume 43(2), pp. 46–52 (2009)
11. Smyth, B., McClave, P.: Similarity vs. Diversity. In: Proc. of the 4th International Conference on Case-Based Reasoning, pp. 347–361 (2001)
12. Ziegler, C.N., McNee, S.M., Konstan, J.A., Lausen, G.: Improving Recommendation Lists Through Topic Diversification. In: Proc. of the 14th international conference on World Wide Web, pp. 22–32 (2005)